A New Growth-regulated Complementary DNA with the Sequence of a Putative Trans-activating Factor

De-Hui Ku, Chung-der Chang, Janice Koniecki, Linda A. Cannizzaro, Leslie Boghosian-Sell, Hansjürgen Alder, and Renato Baserga

Department of Pathology and Fels Institute for Cancer Research and Molecular Biology, Temple University Medical School, Philadelphia, Pennsylvania 19104

Abstract

A new complementary DNA (cDNA) clone has been isolated by differential screening of a cDNA library. The cognate RNA of this clone, called SC1, is growth regulated in human, mouse, and hamster cell lines. Its kinetics of growth regulation (time of increase in mRNA levels, sensitivity to cycloheximide, behavior in G₁-specific temperature-sensitive mutants) classify the SC1 gene as a late growth-regulated gene, like the histone genes and the genes coding for the proteins of the DNA synthesis apparatus. By run-on assay, there is a modest increase in transcriptional rates after serum stimulation, which is not sufficient to explain the sharp increase in mRNA levels. The SC1 gene localizes to human chromosome 6p21–22. In bacteria, the SC1 cDNA clone makes a protein of M, 39,000, in agreement with the putative reading frame. The amino acid sequence derived from the cDNA sequence indicates a previously unknown gene with a domain strongly suggestive of a trans-activating domain. The SC1 gene can be considered as coding for a possible new trans-activating factor that could play an important role in the transcription of genes required for the later stages of cell cycle progression.

Introduction

In recent years, several growth-regulated cDNAs have been isolated by differential screening of cDNA libraries in a number of laboratories (1–4). Although the fact that the expression of a gene is growth regulated does not necessarily mean that the gene is also growth regulatory, the identification and cloning of growth-regulated cDNAs is always of interest since: (a) some of them may actually be growth regulatory, and (b) they are indicative, at worst, of the pattern of gene expression during the cell cycle.

Growth-regulated genes can be divided into two large categories: the early growth-regulated genes, of which c-myc and c-fos are probably the best known prototypes, and the late growth-regulated genes, of which the prototypes are usually genes coding for the proteins of the DNA-synthesizing machinery like thymidine kinase, DNA polymerase-α, PCNA, and others (see “Discussion”). Until recently, early growth-regulated genes were the object of most investigations (see above), but in the past few years, a number of laboratories have also been trying to identify and clone late growth-regulated genes, i.e., genes that are maximally expressed at the G₁-S boundary or during the S phase of the cell cycle. In this paper, we describe the identification and cloning of a cDNA whose cognate mRNA is maximally expressed at the onset of DNA synthesis and during S phase. This clone, which we have designated as SC1, is not only exclusively growth regulated as a G₁-S boundary gene, but it has in its putative amino acid sequence suggestive of a trans-activating factor. It could therefore play an important role in the regulation of expression of other genes that are necessary for the G₁ to S transition or for the entry of cells into G₂ and mitosis.

Results

One of the clones isolated by differential screening of the cDNA library (see “Materials and Methods”) and designated SC1, was found to be growth regulated, as demonstrated in Fig. 1, an RNA blot from BALB/c3T3 cells, which are exclusively growth regulated (5, 6). The SC1 mRNA is not detectable in quiescent 3T3 cells (Lane 3), nor is it detectable when the cells are treated with platelet-poor plasma (Lane 1). The SC1 mRNA becomes detectable at about 8 h after stimulation with serum (Lane 5), reaching a peak at 16–24 h (Lanes 6 and 7), which is also the peak of DNA synthesis in these cells, under these conditions (see also the mRNA levels for histone H3 in the same blot). If the cells are treated with 1 μg/ml of cycloheximide, the induction of SC1 mRNA is totally abolished (Lane 2). On the basis of these results, the mRNA of the SC1 gene is similar in its kinetic behavior to that of other genes that are maximally expressed at the G₁-S boundary, such as thymidine kinase (7–9), PCNA (9, 10), thymidylate synthase (11), RNA primase (12), and DNA polymerase-α (13). The steady-state mRNA levels of these genes reach a maximum at about the time of the onset of DNA synthesis, and all of them are inhibited by relatively low concentrations of cycloheximide.

Isolation and Sequencing of the Human cDNA for SC1

The original SC1 clone was isolated from a 3T3 cDNA library. Judging from the size of the mRNA and the length of the insert, the mouse clone was, in all likelihood, an incomplete cDNA. We then proceeded to the isolation of the corresponding human SC1 clone from an Okayama-Berg library (14). The isolated human cDNA clone contained an insert that was judged to be full length, or nearly full length, since it was approximately 2.6 kb, the same size as the estimated mRNA band on Northern blots, which is between 2.6 and 2.7 kb (a 200-base polyadenylate should be added).
Both the mouse and the human cDNA clones were sequenced. The nucleotide sequences of these two clones are shown in Fig. 2, where the deduced amino acid sequence of the longest open reading frame is also given. In the putative coding sequences (see below), the two clones show extensive similarities, indicating that the gene is quite well conserved between mouse and human. In fact, there are extensive similarities also in the 5′ presumably untranslated region. The murine cDNA stops at the point that corresponds to nucleotide 1179 of the human cDNA clone, which extends, instead, to nucleotide 2473.

A computer analysis of the putative coding sequence revealed that, in all three frames, there are only short stretches that are terminated by termination codons. For instance, the methionine at nucleotide 497 is terminated at nucleotide 662, and the methionine at nucleotide 1078 is terminated at nucleotide 1143. The longest open reading frame was the one given in Fig. 2 which extends from nucleotide 376 (initiation codon) and terminates at nucleotide 1444 (including the termination codon), coding for a putative protein of 359 amino acids. There are also other methionine codons in the 5′ untranslated region, but they all terminate early to give a very small protein. For instance, the methionine at nucleotide 331 is terminated at nucleotide 340, and the methionine at nucleotide 335 is terminated at nucleotide 422.

A computer search failed to reveal significant homologies between the putative sequence of SC1 and other genes or protein sequences in the GeneBank. However, an inspection of the putative sequence revealed that, from amino acid 188 to residue 265, SC1 contains 18 prolines of the 77 residues. This large number of prolines has been considered characteristic of a trans-activating domain (see below). The size of the protein expressed in Escherichia coli also corresponds to this putative sequence (see below).

**Growth Regulation of the SC1 mRNA.** We have shown above that the levels of SC1 mRNA are growth regulated in BALB/c3T3 cells. We have also investigated its growth regulation in other cell lines. Fig. 3 shows the growth regulation of the SC1 mRNA in human diploid fibroblast WI-38 cells. Here, too, the mRNA for SC1 is barely visible in G0 cells but is clearly visible in cells stimulated with serum for 16 or 24 h (Lanes 2 and 3), which, in these cells, also represents the peak of DNA synthesis. The amounts of RNA in each lane were monitored as described in “Materials and Methods.”

Another way of determining the cell cycle behavior of a gene is to investigate it in a G0-specific, temperature-sensitive mutant of the cell cycle. In these cells, it has been clearly established that the early growth-regulated genes, like c-myc, c-fos, and others, are equally induced at the permissive and at the restrictive temperatures (2, 15), whereas genes that are maximally expressed at the G1-S boundary, like histones (16), thymidine kinase, (17), PCNA (9), and DNA polymerase-α (18), are not expressed at the restrictive temperature, although they are normally expressed when the cells are stimulated at the permissive temperature. We have investigated the mRNA levels of SC1 in TK-13 cells (19), which are a TK- derivative of 13 cells (20), which in turn is a G0-specific Is mutant that arrests in G0 at the restrictive temperature. The RNA blot in Fig. 4 shows that the levels of SC1 mRNA increase when cells are stimulated with serum at the permissive temperature (Lane 2), they are not above the G0 levels when the cells are stimulated at the nonpermissive temperature (Lane 3), and again, they are markedly decreased when the cells (Lane 4) are stimulated by serum at 34°C in the presence of cycloheximide. These experiments further confirm that the SC1 mRNA is growth regulated like a late G1-S boundary gene.

In Fig. 1, we have shown that the levels of SC1 mRNA are increased when the cells are stimulated with serum but not when the cells are stimulated with platelet-poor plasma. Platelet-poor plasma is essentially serum without platelet components. We wanted to confirm this hypothesis by stimulating the BALB/c3T3 cells with PDGF only. The results are shown in Fig. 5, where it is apparent that the mRNA levels of SC1 are increased by stimulation with PDGF only, although the increase is not as marked as when the cells are stimulated with serum. This conclusion is based on the comparison of the levels of expression of histone H3, which is clearly higher in Fig. 5 than in Fig. 1. The ratio histone H3:SC1 is considerably higher with PDGF than when the cells are stimulated with serum.

**Run-on Transcription Assay of the SC1 Gene.** The transcriptional activity of the SC1 gene was determined by a run-on transcriptional assay (21, 22) in BALB/c3T3 nuclei. The results are shown in Fig. 6. Lane 1 is from nuclei of serum-deprived cells. Lanes 2 and 3 are from nuclei of cells stimulated for 16 and 24 h, respectively, with 10% serum. When histone H3 is used as a probe, there is a marked increase in transcription rates at 24 h. With the other three genes examined, PCNA, SC1, and β-actin, there is a modest increase above G0, an increase which is not sufficient to explain the marked increase in the levels of SC1 or PCNA mRNA. Thus, although there might be a modest transcriptional component in the increase in SC1 mRNA levels, a posttranscriptional com-
The image contains a section of a scientific document with amino acid sequences and nucleotide sequences. The text appears to be discussing partial sequencing of proteins and nucleotides, possibly related to mouse SC1 and human SC1. The sequences are presented in a tabular format with columns for nucleotides and amino acids, and rows for different sequences. The text is dense and technical, typical of scientific literature. The figure at the bottom, labeled 'Fig. 2', indicates that the image is part of a larger discussion on the comparison of partial mouse SC1 and full-length human SC1 cDNA sequences with the amino acid sequence of the predicted protein. The 'Top row' refers to partial mouse SC1 cDNA sequence, the 'Middle row' to full-length human SC1 cDNA sequence, and the 'Lower row' to the lower untranslated region of the predicted protein.

The polyadenylation signal is underlined. The lower row represents the predicted human SC1 amino acid sequences.
ponent must be invoked to explain the growth regulation of the mRNA levels for this gene.

The Putative SC1 Protein. We were anxious to determine whether the putative open reading frame was indeed correct, and we thought that the best way to demonstrate it was to express this protein in E. coli using our cloned SC1 cDNA. The SC1 cDNA insert was subcloned into a pUC18 vector, which was then expressed in E. coli. The results are shown in Fig. 7, which indicates that the presence of this construct results in the expression of a protein of approximately M. 39,000, which is the size suggested by the open reading frame shown in Fig. 2. We therefore believe that the open reading frame is the correct one and that the SC1 cDNA codes for a protein of approximately M. 39,000, despite the 2.5-kb nucleotide sequence. It should be noted that no other sequence in any reading frame of the human SC1 cDNA would give a polypeptide of comparable size; indeed, the second longest reading frame would give a protein of M. 10,000.

Chromosomal Localization of the SC1 Gene. The chromosomal localization of the human SC1 gene was determined as described in “Materials and Methods.” A total of 150 metaphases were analyzed, and a total of 271 grains were scored over chromosomes (Fig. 8). The majority of grains were situated over the short arm of chromosome 6 (6p) with 78 of 271 or 29% of the total grains hybridized to chromosomes. The primary site of hybridization was in the 6p21–22 region containing 74 of 78 (95%) grains of the total hybridized to 6p (Fig. 9).

Discussion
The search for growth-regulated genes has been the object of several laboratories, since the first ones identified by differential screening of cDNA libraries were reported in 1983 (1; see also Refs. 2–4). The rationale for this search is that the identification of growth-regulated genes serves as the basis for a map of gene expression during the cell cycle and during the transition of cells from a nonproliferative to a proliferative stage. In turn, a genetic map of cell cycle progression could be valuable for our understanding of the mechanism(s) that control cell growth. In the search for growth-regulated genes, the emphasis has usually been on genes that are differentially expressed in the first 2–3 h after quiescent cells are stimulated to proliferate (1–4, 23); genes whose mRNA levels increase sharply in the first few hours after a mitogenic stimulus are referred to as “early growth-regulated genes” (24). Several such genes have been identified, many as a result of large scale differential screening (1–4, 23), and others on an individual basis, yielding a long and varied catalogue of early growth-regulated genes, which include protooncogenes (4, 25, 26), transcription factors (4, 25), growth factors (26, 27), “zinc-finger” proteins (28), etc.

A second set of growth-regulated genes is characterized by the fact that their mRNA levels increase later after mitogenic stimulation and reach a maximum at the time the cells enter S phase. Again, like early growth-regulated genes, they constitute a heterogeneous group, which includes protooncogenes like c-myc (18, 29), core histones (30), and genes coding for proteins involved in the synthesis of cellular DNA, like thymidine kinase (7–9), RNA primase (12), PCNA (9, 10), thymidylate synthase (11), and DNA polymerase-α (13). Late growth-regulated genes are distinguished from early growth-regulated genes by two properties; their expression is extremely sensitive to cycloheximide (8, 9), whereas early growth-regulated genes are insensitive (3, 4), and their expression is also inhibited in G1-specific ts mutants of the cell cycle stimulated at the restrictive temperature (9, 16, 17).

The cDNA clone reported in this paper is obviously a late growth-regulated gene, from the kinetics of appearance of its mRNA (Fig. 1), its sensitivity to cycloheximide (Fig. 1), and its behavior in a G1-specific ts mutant of the cell cycle (Fig. 4). Although its transcription rate increases modestly in serum-stimulated cells (Fig. 6), the increase is not sufficient to explain the sharp increase in mRNA levels. In this respect, SC1 behaves again like other late growth-regulated genes, notably PCNA (9), thymidine kinase (9), and core histones (16), in which the mRNA levels have a strong component of posttranscriptional regulation. The human cDNA clone is full length or nearly full length. It has 2473 bases; since the size of the mRNA band on Northern blots is less than 2.7 kb (with the poly(A) tail), our cDNA must be full length or, at the most, miss a few nucleotides. The 5′ to 3′ direction is out of the question, since this is a clone from the Okayama-Berg library. The 3′ untranslated region is unusually long, but long untranslated 3′ regions have already been reported. Furthermore, the only polyadenylation signal in our sequence is at residue 2437.
Fig. 5: RNA blot of BALB/c 3T3 cells stimulated with PDGF (1 ng/ml). Same methods as in Figs. 1 and 3. Lanes 2, 3, and 4, cells stimulated with PDGF for 4, 8, 16, and 24 h, respectively. Both SC1 and histone H1 probes were used.

Fig. 6: Run on transcription in BALB/c 3T3 nuclei. Lane 1, nuclei from deprived cells; Lanes 2 and 3, nuclei from cells stimulated for 16 and 24 h, respectively, with 10% serum. The following probes were used: histone H1 (16), full-length pCNA cDNA (17), SC1, lambda phage negative control, and ß-actin (17).

We sequenced both strands, and some points (especially termination codons) were sequenced repeatedly; we feel confident, therefore, that the putative open reading frame is the correct one.

Expression of SC1 in E. coli has confirmed the open reading frame and putative amino acid sequence, deduced from the nucleotide sequence in Fig. 2. It gives a protein of approximately M, 39,000, as predicted by the amino acid sequence in Fig. 2; no other sequence in any reading frame would code for a polypeptide that even remotely approached the size of M, 39,000.

Fig. 7: Expression of SC1-Lac Z fusion protein in E. coli system. Lane 1, protein isolated from E. coli containing plasmid pSC1 without IPTG stimulation; Lane 2, same, except stimulated with 10 mM IPTG; Lane 3, protein molecular weight standard; Lane 4, same as Lane 1, except stimulated with 1 mM IPTG.

Neither the nucleotide sequence nor the putative amino acid sequence has significant similarities to any gene or protein whose sequences are present (at the time of writing) in the GeneBank. However, from amino acid 188 to amino acid 265, the SC1 gene product contains 18 prolines, or 24% proline content. This is characteristic of certain trans-activators (31), of which the prototype is CTF/NF-1 (32). According to Mitchell and Tjian (31), proline-rich regions are also found in other mammalian transcription factors, such as AP-2, jun, and OCT-2. The SC1 putative sequence, like CTF/NF-1, has doublets and triplets of proline in the general proline-rich region. Clearly, the ability of SC1 to function as a transcription factor will have to be determined by experimentation. Temporarily, we can say that SC1 is a late growth-regulated gene that is a good candidate for a novel transcription factor.

Finally, the SC1 gene maps to the short arm of human chromosome 6, in the region 6p21-22. Within this same
segment of chromosome 6, several other genes have been mapped. These include the loci which code for the human major histocompatibility complex, tumor necrosis factor, the homologues of the PIM oncogene, as well as type Xla2 collagen (33). The organizational relationship of each of these genes with SC1 is not yet known. However, further investigations in this area may yield interesting clues as to whether SC1 has any regulatory role in their expression.

Materials and Methods

Northern Blots. Total RNA was extracted from different cell lines by the method of Chomczynski and Sacchi (34), and RNA blots were carried out by standard procedures (35). A 1.5-kb EcoRI fragment isolated from mouse SC1 cDNA was labeled by the random primer method (36) and was used as a probe. The RNA amounts in each lane were monitored by hybridization to a probe that is constantly expressed throughout the cell cycle (9).

Run-on Transcription. Run-on transcription was carried out exactly as described by Groudine et al. (21). A 1.5-kb AfflII human SC1 probe was used in this experiment. The other probes are described in the paper by Chang et al. (37).

Screening of a cDNA Library. A Lambda Zap II cDNA library (Stratagene) was prepared from mRNA extracted from BALB/c 3T3 cells stimulated with serum for 16 h. The library had in excess of 10^7 individual recombinants. The library was then differentially screened by the usual procedure; namely, replica plates were hybridized to single stranded cDNA from two different sources: (a) cells stimulated for 16 h with serum, as in the preparation of the library, and (b) cells stimulated for 16 h with serum after addition, at 8 h, of 1 μg/ml of cycloheximide. A total of 500,000 recombinants were screened.

Screening of Okayama-Berg Library. A total of 240,000 recombinants were screened according to the method previously described (2). The 1.5-kb EcoRI fragment of mouse SC1 clone was used as a probe after labeling by the random primer method (United States Biochemical Corp.). One positive clone was isolated with a 2.4-kb insert.

Sequence Analysis. Sixteen synthesized primers were used to directly sequence the human SC1 Okayama-Berg cDNA clone in both strands by Sequenase (United States Biochemical Corp.). A 1.5-kb EcoRI fragment of mouse SC1 cDNA clone was subcloned into pGEM3Z vector and sequenced using T7 and SP6 primers, plus six synthesized primers. Sequencing by standard procedures was carried out on both strands, with overduplications and frequent repetitions (38, 39).

Plasmid Construction. A plasmid pUCSC1 was constructed by using polymerase chain reaction techniques such that the presumed open reading frame of SC1 was in the same frame as β-galactosidase in the pUC18 plasmid vector. The plasmid was then checked by sequence analysis to make sure that no mutation had occurred during polymerase amplification.
E. coli System. Total protein was isolated from E. coli (TB1) strain containing the pUCSC1 plasmid according to the method of Som4 after stimulation with 10 mm IPTG or without IPTG. The proteins were then run on sodium dodecyl sulfate-polyacrylamide gel and stained with Coomassie blue according to standard procedures (Promega).

Methods for Chromosomal Localization. The probe was labeled with [3H]dATP to a specific activity of 2.4 × 108 cpm/μg DNA. A final probe concentration of 0.07 μg/ml was added to each slide. Metaphase chromosomes were obtained by culturing peripheral blood lymphocytes from a normal (46,XY) male for 72 h in RPMI medium supplemented with 20% fetal bovine serum. Cell cultures were harvested according to standard procedures.

In situ hybridization was performed using a modification of the standard protocol (40, 41). Slides containing chromosome preparations were aged first at 4°C for 7-14 days and then treated with ribonuclease A (Sigma) for 1 h at 37°C. Chromosomal DNA was denatured at 70°C for 2 min in a 70% formamide-2× standard saline citrate mixture (pH 7.0). Probe DNA was denatured in a hybridization mixture of 50% formamide, 2× standard saline citrate, and 10% dextran sulfate (pH 7.0). Hybridization was carried out at 37°C for 16 h. Slides were rinsed the next day at 39°C, dehydrated in an ethanol series, and air dried. Slides were then dipped in nuclear track emulsion (Kodak NTB-2), air dried, and stored in light-tight boxes at 4°C. At different time intervals, slides were developed and fixed at 15°C, air dried, and then stained with a modified Wright's Giemsa staining protocol (41). Each slide was dipped 10-16 s in a pH 9.2 borate buffer (50 mm NaSO4-2.5 mm Na2B4O7) at 35°C and stained 2-5 min in a staining dish containing 2-3 parts pH 9.2 buffer:1 part Wright's Giemsa stain (T. Baker & Co). Single grains situated on non-overlapping chromosome regions were counted and scored.

Acknowledgments

We acknowledge the technical assistance of Graziella Zani.

References


*4 T. Som, personal communication.